UCLA Anderson | City of LA Datathon Sponsored by Accenture

Problem Statement & Instructions

Many possible insights can be found in the City's massive payroll data (which comprise over 299,000 observations and 35 attributes), but the City is most interested in understanding the historical payroll data for each department and job title. Help them to get a better picture of the realities of payroll in Los Angeles, by uncovering insights in the data and provided a descriptive overview. The City will also be evaluating the quantitative accuracy of your predictions by asking teams to predict the actual 2018 payroll data which has been removed in advance from the data set.

Below you will find a link to the data you will need in this competition. You will have information about past expenditures for individuals from years 2013 to 2017 along with their department names and job titles. The City has provided data for full time, part time, and employees hired for events. Moreover, hourly base rate and total overtime expenses are captured in the dataset.

Given the information, the City of Los Angeles wants you to predict the Q1 and Q2 2018 payments for each employee. This will take some data cleaning and will challenge you to use advanced methods to predict with accuracy. Furthermore, you are asked to create an optimization plan and recommendations that can help the city to reduce the inefficiencies in the system. Your efforts will result in optimized overtime expenses, which in turn will make the City more efficient.

URL to download the data dictionary and the data set:

https://controllerdata.lacity.org/Payroll/City-Employee-Payroll/pazn-qyym/

Deliverables from teams due on November 11th, 2018 by 11:59 p.m.

- Document describing the improvement they suggest to the city of LA based on this data-set (1,000 words max)
- payroll_2018_FirstName_LastName_Team#.csv file emailed to Paul Brandano at paul.brandano@anderson.ucla.edu

  (with FirstName, LastName, Team # of any of the team members)
- R or Python code

Evaluation Criteria for finalists' selections

The scoring will be based on the below two criteria:

1. Qualitative evaluation (50% weight) – no more than 2 pages of summary findings and recommendations
2. Model evaluation (50% weight):

- The total number of matches within the nearest 1000th decimal place.
  (For example, for an *actual* payments value of 10,000, your prediction needs to lie between 9,500.00 to 10,499.99 to be considered a match.)
- RMSE score (lower the better) will be used only in case of a tie.

4 Finalists will be declared on November 13th, 2018

Deliverables from the Finalists due by November 16th, 2018 by 10:00 a.m.

- A PowerPoint presentation of analysis, findings and recommendations
- No slide limit; should be presented within 10 minutes + 5 minutes of Q&A

November 16th, 2018 Finalists present findings to the judges in Korn Hall at 3:45 p.m.

- Teams will have 10 minutes to present
- Judges will have 5 minutes for clarifying questions